ISSN No:-2456-2165

Crime Analysis and Prediction using Machine Learning

[1]Nazma Sultana Shaik, [2]K. Sai Krishna, [3]P. Naveen Kumar
[1,2,3]Department of Information Technology
[1,2,3]Vignan's foundation for science Technology and Research

Abstract:- Crime is one in every of the most important hurdle in today's global and it's growing at the hearth place tempo that is a primary motive for concern. There is a want to screen and maintain music of all the crimes in order that it may be utilized by police branch to analyze the instances without problems and quickly. In this experiment, a gadget studying K-way set of rules is used to are expecting and examine the crime with inside the metropolis of Chicago. The dataset of crimes in Chicago is to be had on the kaggle website, this is used because the dataset to make prediction and to visualize the styles and developments of various crimes. The different cause of this mission is to assess how lots okay – way set of rules is viable to decide and clear up the present day problem.

Keywords:- *Crime Analysis, Training Datasets, Decision tree, Naïve Bayes, k Nearest Neighbor (KNN).*

I. INTRODUCTION

The rate of crime is rising on a daily basis as current technologies and high-tech ways assist criminals in carrying out their unlawful activities. According to the Crime Record Bureau, burglary, arson, and other crimes have increased, while murder, sex, abuse, gang rap, and other crimes have increased [2]. Data about crime will be gathered from a variety of blogs, news outlets, and websites. The massive data is used to create a crime report database as a record. The knowledge gained via data mining techniques will aid in the reduction of crime by assisting in the speedier identification of criminals as well as the areas most affected by crime.

The expansion of research methodologies are aiming to get the resources in background from the crime reports that are available in order to get a better view about criminal behavior and moreover it is used to prevent future crimes. It has resulted in the rapid growth of the crime data records that are combined along with the data analytics. In this way crime can bea complicated social problem that has grown in response to major key elemental developments. The elements that lead to a rise in criminal tendencies must be discovered by law enforcement agencies. There is always a need for crime prevention techniques and policies to combat this.

The two major law enforcement professions like Data mining as well as artificial intelligence, simply called as DM & AI techniques are becoming as a result of technological advancements, research, and information.

In order to predict and prevent the crime we have to do the crime analysis by using the Data Mining techniques. Law enforcement organizations deal with a lot of data. Valuable information can be obtained by the processing of that large amount data. The criminal data can be processed by using the several models offered by the law enforcement agencies in the perspective of preventing the crime.

II. CRIMEDATAANALYSIS

Secure agencies require the collection and analysis of crime-related data. The most essential factors that must be addressed are the employment of coherent methods to classify this data based on the rate and place of incidences, identification of the underlying pattern among the broken laws at different times, and forecasting of their future relationship. Hot spot analysis is one of the most popular methods. Point method of analysis and clustering/distance statistics are two of the most used methodologies for this purpose. The finding of patterns or trends using techniques like as data mining, text mining, spatial, and self-organizing maps is another prominent approach.

Crime analysis is a branch of criminology tasked with investigating and uncovering crime and its links to criminals. The goal of law enforcement is to identify the features of crime. The initial step in generating further analysis is to identify crime characteristics. Because of the large volume of crime data and the intricacy of the links between them, criminology is an ideal topic for data mining.

The primary goal of crime analysis is to:

- Data mining can be used to investigate very enormous datasets with a large number of variables that are beyond the scope of a single analyst, or even an analytical team or task force.
- Based on the leads and circumstances accessible to law enforcement agencies, the goal of criminal investigative analysis is to find parallels in specific occurrences of serial crimes.
- Tactical, operational, and strategic levels of crime analysis are all possible. As rapidly as possible, crime analysts examine crime data, arrest reports, and police calls for service in order to spot emerging patterns, series, and trends.

ISSN No:-2456-2165

A. CRIMEANALYSISMETHODOLOGY

The several methodologies that help in the evaluation of crime are:-

- Data Collection
- Classification
- Pattern Identification
- Prediction
- Visualization

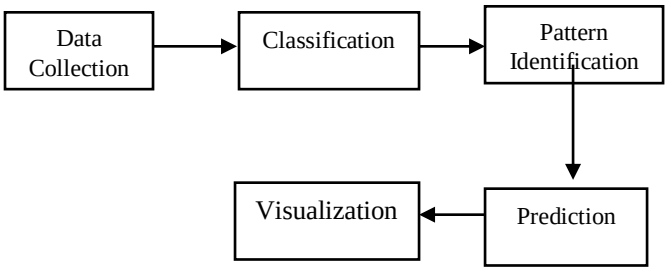

Fig. 1: Crime Analysis Steps

B. Data Collection:

The statistics collection is first approach in crime analysis. Data's are accumulated from numerous one-of-a-type websites, records on web, web sites and blogs. The accumulated statistics is stored into database for further process. This is unstructured and iseasy to use and flexible because it is made up of object oriented programming. The data format of the crime statistics is mostly unstructured. It is clearly observed by comparing the number of fields and content storage type in those fields from one format at one place to the different format at another place. To overcome that we have to use a schema that will operate more efficiently on small datasets like time to time.in accordance with that the complexity will be reduced by removing the joins. Along with these there are many other uses comes with unstructured data are:

- High amount of all the three types of data
- Programming language that is simple to use and flexible in type.

C. Classification:

In this step use Naive Bayes Algorithm that is supervised getting to know method. One of the probabilistic classifier namely Naïve Bayes to which whenwe given an enter offers a best chance to spread the data evenly to all the training sets as opposed to imparting a unmarried output. One of the primary benefits other than simple in the classification of Naïve Bayes is insurance and when compared to the other it is faster in the classification. Compare two different set of rules like SVM which in terms called as Support Vector machine which takes plenty of memory. Using naïve Bays set of rules is create a version via way of means of education crime records associated with murder, vandalism, intercourse abuse, burglary robbery, rape attacks etc. The main purpose of the naïve Bayes is to work properly over a small amount of data that is having a limited number of fields which contains data regarding the crime. Estimating chance on occasion whilst checking a chance **P[A]*P[B/D]*P[C/D]*P[E/D]** where in **P[C/D] =0[2]**.

D. Pattern Identification:

A 1/3 step is the sample identity wherein we've pick out developments and styles in crime. For locating crime sample that takes place often we're the usage of apriority algorithm. Apriori may be used to decide affiliation rule which spotlight fashionable developments with inside the database. By the usage of sample identity it's going to allows to the police officers in an powerful way and keep away from the crime occurrences mainly vicinity with the aid of using supplying security, CCTV, solving alarms etc.

E. Crime Prediction:

The 2d Approach is predicting the crime kind that could arise in a particular vicinity inside unique time. To expect an anticipated crime kind is offer 4 associated functions of the crime. The functions are: prevalence month, the prevalence day of the week, the occurrences time and the crime vicinity. Prediction is pointing out opportunity of an occasion in destiny length time. A Classification technique is used crime prediction in facts mining classify regions into hotspots and bloodless spots and to predictive a place might be a hotspot for residential burglary. Variety of class strategies are used for predicting the crime:-

- KNN
- J48
- SVM
- Neural Networks
- Naïve Bayes and ensemble learning

Linear Regression strategies also are used for predicting the crime prediction. Based at the crime probability. The system for a regression line is

Y=aX + b where, Y is the expected score, b is the slope of the line, and A is the Y intercept. b = r sx\sy

And the intercept (A) may be calculated as A=MY –bMX.
Some Theories used to predicting the crimes are:

- Integrated theory
- Biological theory
- Psychological theory
- Sociological theory
- Conflict theory
- Victimization theory
- Choice theory

F. Visualization:

Visualization can be done by using the heat map which shows the activity in several areas, whereas the dark areas represent the Low Activity and the Light colors indicates the areas where activity is very high.

Advantages of using heat map are:-

- Colored images based on both numerical and category.
- Choice to select the data we want to analyze.
- Unrelavent data will not be taken into consideration.

ISSN No:-2456-2165

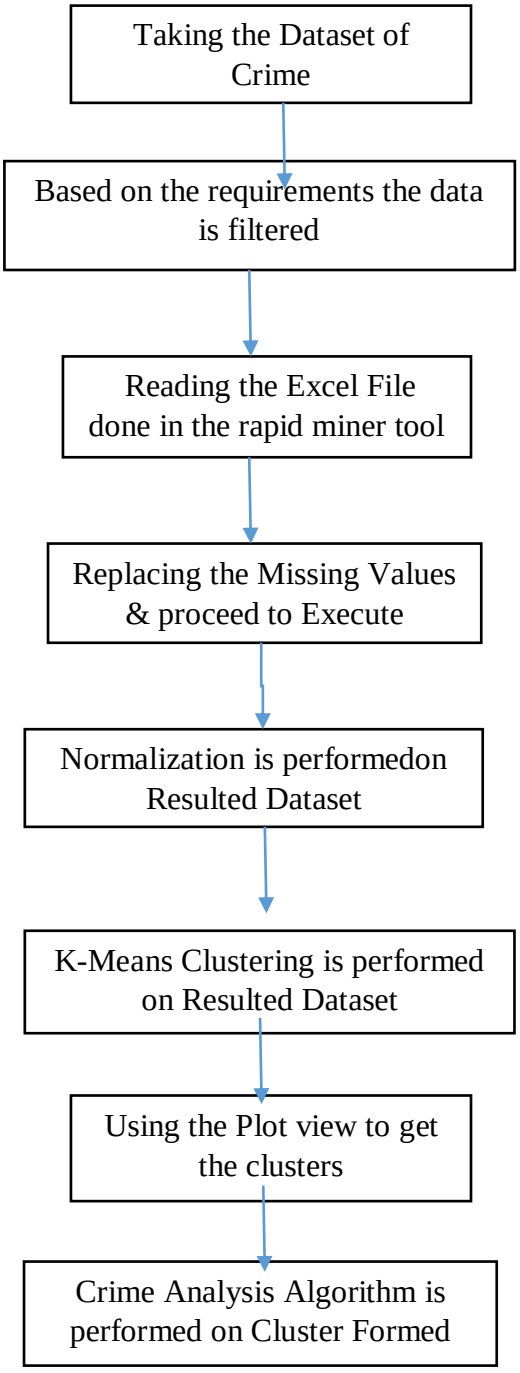

Fig. 3: FLOWCHART OF CRIMEANALYSIS

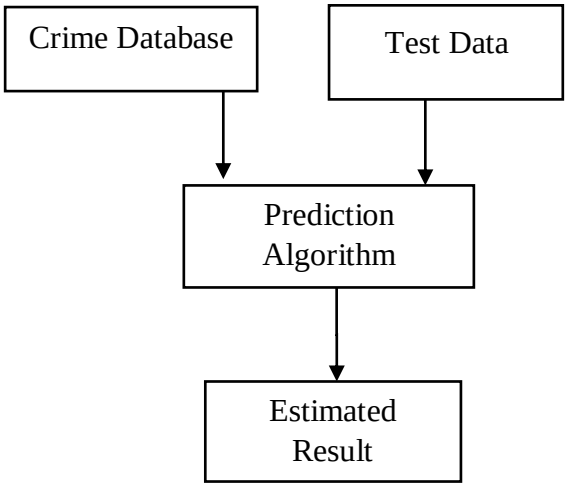

Fig. 4: PREDICTING FUTURE CRIME TRENDS

III. DATA

The Austin plice department write a reply to all the crime incidents that come into notice to the police and store those replies in a dataset. The data stored in that manner is presented from the year 2003 to till now. The dataset is going to update every week with the new incidents that are reported in that week. By this we will get to know the how much of the data is going to be collected in a week. Based on the information gathering technique, which is different from one to other the data collected is also in different formats according the nature of the information collector. The data uploaded in the dataset is to be coming from these sources, so it is likely to be different when compared to the same dates by different in gathering the information source.

Now a days each report generated by the police is having such a unique number which is used to know all the details about the incident along with the time date and the person who take responcibity of that particular incident. Which makes it easy for any of the police in any other place to know the information using that reference number in the police database. Which makes the police department a benefit of no longer storing the files written at the time of the incident reporting.

Now a days the legal responsibility to those papers which are written and stored by the Austin police is not provided. Since storing such amount of papers in the police station is also a big challenge that should we take into consideration. Using the codes in the law namely Segment 552. 301© states that the public can be given freedom to talk about the incident with the police officers by viewing the data that has been sent via E-Mail In the dataset includes distinctive kinds of crimes (attributes) are taken into consideration like murder, rape, kidnapping, dacoit, robbery, burglary, cheating, dowry deaths, arson etc.

ISSN No:-2456-2165

S.No	Name Of The Crime	Records Found
1	Violentcrime	20384
2	Burglary	21048
3	Vehiclecrime	17964
4	Anti-socialbehavior	46152
5	Robbery	7452
6	Drugs	13425
7	Theft from person	1486
8	Other theft	8945

Table 1: Different crime data information

ISSN No:-2456-2165

IV. ALGORITHMS

The Algorithms used in our experiment are

- Instance based algorithm
- Decision tree
- Linear regression
- K-means algorithm

A. Instance Based Algorithm

The example primarily based totally set of rules is likewise known as the gadget primarily based totally getting to know is a own circle of relatives of getting to know set of rules that, in preference to appearing specific generalization, compares new issues times with example visible in schooling, that have been saved in memory. These saved their schooling set while predicting a cost or magnificence for a brand new times, they compute distance schooling times to make a decision. The blessings of the Instances primarily based totally Algorithm is it over different strategies of system mastering is its capacity to evolve its version of system mastering is its capacity to evolve its version to formerly unseen data. Instance primarily based totally beginners can also additionally really shop a brand new example or throw an antique example away. The Disadvantages of the times primarily based totally Algorithm are its want greater garage and computational complexity.

B. Linear Regression

It is easy shape of regression. Linear regression tries to version the connection among the 2 variables with the aid of using becoming a linear equation to look at the data. That is extensively utilized in statistics. The unknown parameter i.e., weight of the impartial variables, are predicted from the education data for the linear capabilities which is used for this purpose. This may be used to are expecting the values One of the maximum not unusual place estimating technique is least imply squareSimple regression, multiple regression, and rhythmic regression are the one of the different models in the Algorithm of linear regression. Which are suitable for high-dimensional data and only accept nominal binary attributes [1]. The main benefit of linear regressions is to better understand the variables that could affect your success in the weeks, months and years to come. Linearity is the main setback for the regression models. If the data is nonlinearly dependent, the most adequate and suitable line can be implemented by using the linear regression, which may not fit very well.

C. Decision Tree

The decision tree is used in several ways like predicting and classifying data. We have to learn a function inorder to do the classification, that function is called intervals defined by distribution of the values belonging to the individual attributes. Advantages of the choice timber are It is quite simple to recognize and assist decide worst, great and anticipated values for unique scenarios. It is able to be blended with different choice techniques. Some of the Disadvantages of the Decision tree are they are unstable, they are regularly exceedingly inaccurate, Calculation can get very complex.

Area Sensitivity	YES	YES	NO	YES	YES	NO
Notable Event	YES	YES	NO	NO	YES	YES
VIP Presence	YES	NO	NO	NO	YES	NO
Criminal Group	NO	YES	YES	NO	YES	NO
Crime	YES	NO	NO	NO	YES	NO

Table 2: Describing the Attributes

D. K-Means algorithm

K –method is also very easy method to handle the datasets and maximum normally used portioning set of rules a number of the clustering set of rules in medical and commercial software. Acceptance of okay method is particularly because of its being easy .This set of rules is likewise used in a appropriate way to make clusters of the large datasets because it has tons much less complexity in the field of computation that grows continuously via way of means of growing of the records points. Advantages of the okay-method set of rules are particularly easy to implement, Scales to massive dataset, Guarantees convergence, effortlessly adapts to new examples. Disadvantages of the okay-method set of rules are choosing manually, being depending on preliminary values, clustering records of various sizes and density.

V. CONCLUSION

As we seen that this paper is targeted on constructing a way to predict the fashions of crime that are occurring very frequently with the experience in the crime consistent within a month. The crime fees in India are growing every day because of many elements which includes boom in poverty, implementation, corruption, etc. The proposed version may be very beneficial for each the investigating organizations and the police reputable in taking vital steps to lessen crime.

The task allows the crime evaluation to evaluation those crime networks by way of numerous interactive visualization. Future enhancement of this studies paintings on education bots to expect the crime inclined regions through the use of system studying techniques. Since, system studying is much like records mining superior idea of system studying may be used for higher prediction. The records privacy, reliability, accuracy may be progressed for better prediction.

REFERENCES

[1.] Review on crime Analysis and Prediction Using Data Mining Techniques, International Journal of Innovative Research in Science Engineering and technology, 2018 by M. Sreedevi , A. Harsha vardhan, Ch.Venkata sai Krishna.

[2.] Review on Crime Analysis and prediction using data mining techniques, International Journal of recent trendsinengineering&research,2019 by Rajkumar and Sakkarai.

[3.] Hitesh Kumar, Bhavana, Ginika's review on Crime Prediction & Monitoring Framework Based on Spatial Analysis, ICCIDS, 2018.